

基于层次贝叶斯模型的成渝地区双城经济圈雾霾监测影响因素分析

骆琳¹, 杨浩蓝², 李勇³

1. 四川旅游学院 大数据与统计学院, 成都 610100

2. 招银网络科技, 成都 610095

3. 成都信息工程大学 统计学院, 成都 610103

摘要:目的 建立科学有效的雾霾影响因素分析方法, 解决成渝地区区域性中轻度大气污染问题。方法 获取成渝地区双城经济圈2014—2022年空气质量逐时监测数据, 结合数据的嵌套结构特征, 构建年度时间-站点-地区三水平层次贝叶斯发展模型, 并进行经典OLS、IGLS估计, 经验贝叶斯和完全贝叶斯的对比分析, 论证了完全贝叶斯方法的优势。结果 $PM_{2.5}$ 浓度受大气污染物、气象、人口、经济、产业结构、农业生产综合因素影响, $PM_{2.5}$ 浓度变化速率受到CO、 SO_2 和城镇化率的影响。结论 利用层次模型建模分析具有嵌套多层次结构的雾霾监测数据, 更科学合理, 再借助贝叶斯统计具有利用先验信息和学习机理的优势, 更有助于提高模型参数估计和预测精度。

关键词:层次贝叶斯; 成渝地区双城经济圈; 雾霾影响因素

中图分类号: O212 文献标识码: A doi: 10.16055/j.issn.1672-058X.2024.0004.014

Analysis of Influencing Factors of Haze Monitoring in Chengdu-Chongqing Economic Circle Based on Hierarchical Bayesian Model

LUO Lin¹, YANG Haolan², LI Yong³

1. School of Big Data and Statistics, Sichuan Tourism University, Chengdu 610100, China

2. CMB Network Technology, Chengdu 610095, China

3. School of Statistics, Chengdu University of Information Technology, Chengdu 610103, China

Abstract: Objective A scientific and effective method for analyzing haze influencing factors is established to solve the regional moderate and mild air pollution in the Chengdu-Chongqing area. **Methods** The hourly monitoring data of air quality from 2014 to 2022 in Chengdu-Chongqing Economic Circle were obtained, and the nested structure characteristics of the monitoring data were used to construct a three-level Bayesian development model at the annual time-site-region level, and the comparative analysis of classical OLS, IGLS estimation, empirical Bayesian and complete Bayesian was carried out, which demonstrated the advantages of the complete Bayesian method. **Results** The $PM_{2.5}$ concentration is affected by comprehensive factors including air pollutants, meteorology, population, economy, industrial structure, and agricultural production, and the change rate of $PM_{2.5}$ concentration is affected by CO, SO_2 , and urbanization rate. **Conclusion** The use of hierarchical modeling to analyze haze monitoring data with nested multi-level structures is more

收稿日期: 2023-02-17 修回日期: 2023-04-04 文章编号: 1672-058X(2024)04-0112-09

基金项目: 国家社会科学基金重大项目(21&ZD153); 四川省社科规划项目(SC21TJ004); 重庆市社会科学规划项目(2020QNJY59; 2020ZDTJ08); 重庆高等教育教学改革研究重大项目(201022); 重庆市教育科学“十三五”规划课题(2020-GX-294); 成都师范学院校级项目(CS22QN01)。

作者简介: 骆琳(1997—), 女, 重庆人, 硕士, 从事经济社会统计与智能社会治理研究。

通讯作者: 李勇(1970—), 男, 重庆人, 教授, 从事智能大数据统计与智能社会治理研究。Email: 397795585@qq.com.

引用格式: 骆琳, 杨浩蓝, 李勇. 基于层次贝叶斯模型的成渝地区双城经济圈雾霾监测影响因素分析[J]. 重庆工商大学学报(自然科学版), 2024, 41(4): 112—120.

LUO Lin, YANG Haolan, LI Yong. Analysis of influencing factors of haze monitoring in Chengdu-Chongqing Economic Circle based on hierarchical Bayesian model[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(4): 112—120.

scientific and reasonable. By incorporating Bayesian statistics, which allows for the utilization of prior information and learning mechanisms, it can further enhance the accuracy of model parameter estimation and prediction.

Keywords: hierarchical Bayes; Chengdu-Chongqing Economic Circle; haze influencing factors

1 引言

2021年10月,国务院印发《成渝地区双城经济圈建设规划纲要》,将成渝地区双城经济圈的建设上升为国家战略。2022年2月,生态环境部编制《成渝地区双城经济圈生态环境保护规划》,明确指出:“十四五”时期易发生区域性中轻度大气污染,进一步改善难度较大。持续深入打好蓝天保卫战,推进大气污染防治又一次作为二十大的重要议题。雾霾影响因素的分析是解决大气污染问题的关键环节,为此,对成渝地区双城经济圈雾霾监测影响因素进行深入研究,有助于推动成渝地区大气环境质量持续改善,建设“高品质生活宜居地”和“美丽中国先行区”。

针对雾霾污染及其影响因素的研究,已有很多学者得出温度、降水、风速、风向、相对湿度、气压与雾霾污染的形成紧密相关^[1]的结论。除自然因素外,也有不少学者从社会经济角度,探讨了人口发展、经济发展、产业结构、能源消耗、交通运输、城市建设对雾霾污染的影响^[4]。雾霾影响因素研究方法方面,学者们一是借助空间统计建模分析:任毅等^[7]基于STIRPAT模型对长江经济带雾霾污染的空间效应与影响因素进行分析;肖翠翠等^[8]运用AR模型,分析辽中南城市群雾霾特征,运用格兰杰因果检验和ISM模型讨论城市间雾霾污染的空间溢出效应。二是基于经典统计建模分析:盛小星等^[9]基于集对分析法对长三角雾霾风险进行评估;洪宇^[10]利用结构方程模型评估各省雾霾水平;王霞^[11]构建改进的灰色关联雾霾影响因素模型,实证分析城市发展、大气污染、气候等因素对雾霾污染的影响。三是利用机器学习算法建模分析:孙亚男等^[12]构建决策树递归分析机器学习模型,量化分析雾霾污染和治理因素的交互影响;沈晓晶等^[13]构建MODIS遥感图像下的雾霾识别神经网络模型,实现对近海上空雾霾监测。

综合以上文献,可发现:指标选取上,现有文献对人为排放中农业面源污染,如化肥、农药施用量少有考虑;研究对象上,主要集中于京津冀、珠三角和长三角这三大经济增长极的雾霾形成原因,将中国“第四增长极”成渝地区双城经济圈作为一个整体系统探讨大气环境和雾霾监测的研究相对欠缺;研究方法上,大多侧

重于传统统计建模和常规机器学习算法。事实上,雾霾监测数据是具有层次结构的多水平数据,监测站点数据会受到所在城市区域经济社会指标的嵌套结构影响,传统统计建模技术难以满足这样的建模条件,建模效果不甚理想。利用具有嵌套多层次结构数据的层次模型进行建模分析,更科学合理。但层次模型的研究主要集中在频率派研究上,极少将贝叶斯思想引入层次模型。其实,频率学派的思路,如采用最小二乘或极大似然法来估计回归模型系数的方法,更适用于数据比较丰富,但对事物的先验认识不足的情况。而贝叶斯方法,更适合数据匮乏或质量不高,对事物的认识较为深刻,能够提出一个较好的先验分布情况。层次模型的结构本身就隐含着贝叶斯统计思想,其低层回归模型的回归系数参数,都视为随机变量,在高一层中展现,这正与贝叶斯核心思想(参数视为随机变量)高度吻合,即层次模型与贝叶斯分析的融合具有内在合理性。

鉴于此,本文获取成渝地区双城经济圈2015—2022年空气质量逐时监测数据,结合监测数据的嵌套结构特征,构建年度时间-站点-地区雾霾监测影响因素三水平贝叶斯发展模型,再基于数据基本分析和模型结果探讨成渝地区雾霾治理存在的主要问题,并提出对策建议。

2 层次线性模型的贝叶斯分析

2.1 贝叶斯估计

贝叶斯估计主要包括经验贝叶斯估计和完全贝叶斯估计方法。经验贝叶斯估计是模型中仅对部分未知参数设置先验分布,而完全贝叶斯要求对模型中所有未知参数都要设置一个先验分布,由此计算出后验分布,所有的统计推断基于后验分布。实质上,经验贝叶斯估计是介于频率派和完全贝叶斯方法之间的一种方法。

层次线性模型的完全贝叶斯方法,能够解决极大似然估计(Maximum Likelihood, ML)和限制极大似然估计(Restricted Maximum Likelihood, RMEL)方法都无法解决的兴趣参数统计推断易受其他辅助参数不确定性的干扰问题,能够确保在对模型每一个未知参数进行统计推断时,完全可以兼顾其他所有参数的不确定性。其基本思路:首先,确定层1回归系数 β 的先验分布,由

于 β 在层次模型中被视为随机变量,其概率性质已经被层 2 模型 $\beta = W\gamma + \delta, \delta \sim N(0, T)$ 所表现,因此,层 2 模型本身就可以视为层 1 回归随机系数 β 的先验分布,就有了层次线性模型第一阶段的先验分布 $p_1(\beta | \gamma, T)$;其次,再设置其他未知参数 $(\gamma, \sigma_\varepsilon^2, T)$ 的先验分布 $p_2(\gamma, \sigma_\varepsilon^2, T)$;再结合已知数据信息 Y ,通过贝叶斯定理,求出所有参数的联合后验分布:

$$p(\beta, \gamma, \sigma_\varepsilon^2, T | Y) = \frac{f(Y | \beta, \sigma_\varepsilon^2) p_1(\beta | \gamma, T) p_2(\gamma, \sigma_\varepsilon^2, T)}{\iiint f(Y | \beta, \sigma_\varepsilon^2) p_1(\beta | \gamma, T) p_2(\gamma, \sigma_\varepsilon^2, T) \partial \beta \partial \gamma \partial \sigma_\varepsilon^2 \partial T}$$

对兴趣参数的推断,可以通过 $f(\beta, \gamma, \sigma_\varepsilon^2, T | Y) = f(Y | \beta, \sigma_\varepsilon^2) p_1(\beta | \gamma, T) p_2(\gamma, \sigma_\varepsilon^2, T)$ 对其他辅助参数求积分所得。若兴趣参数是层 1 随机效应系数 β ,则有

$$f(\beta | Y) = \iiint f(\beta, \gamma, \sigma_\varepsilon^2, T | Y) \partial \gamma \partial \sigma_\varepsilon^2 \partial T$$

同理可对层 2 固定效应系数 γ 、层 1 残差方差 σ_ε^2 、随机系数的方差协方差矩阵 T 等进行参数估计。

2.2 贝叶斯假设检验

层次线性模型的贝叶斯假设检验,主要指利用完全贝叶斯估计方法,对模型中的所有未知参数都给定其先验分布,结合已知数据信息,利用贝叶斯定理,求得所有参数的后验分布,再根据后验分布对模型的参数进行假设检验分析。在两水平层次线性模型的假设检验中,主要涉及三种参数:层 1 回归系数 β 、层 2 固定效应回归系数 γ 和层 2 随机效应的方差协方差成分 T 。

对于贝叶斯假设检验的一般问题,设原假设 H_0 与备择假设 H_1 ,有 $H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1$,其中, $\Theta_0 \cup \Theta_1 = \Theta$ 。若根据数据所得后验分布 $\pi(\theta | x)$,可直接计算两个假设 H_0 和 H_1 的后验概率 α_0 和 α_1 :

$$\alpha_0 = P(\Theta_0 | x), \alpha_1 = P(\Theta_1 | x)$$

再通过后验机会比 α_0/α_1 进行假设检验。当 $\alpha_0/\alpha_1 > 1$ 时,接受原假设 H_0 ;反之,拒绝原假设。当 $\alpha_0/\alpha_1 \approx 1$ 时,可进一步抽样或搜集先验信息再作判定。若引入先验概率,设两个假设 H_0 和 H_1 的先验概率分别为 π_0 和 π_1 ,则根据后验机会比 α_0/α_1 与先验机会比及 π_0/π_1 ,定义 $B^\pi(x) = \alpha_0 \pi_1 / \alpha_1 \pi_0$ 为支持原假设 H_0 的贝叶斯因子 $B^\pi(x)$, $B^\pi(x) \rightarrow$ 大,则支持原假设 H_0 的程度越高。

在两水平层次线性模型中,若要检验层 2 固定效应 γ ,设原假设 H_0 和备择假设 H_1 为 $H_0: \gamma_{kl} = 0 \leftrightarrow H_1: \gamma_{kl} \neq 0$,层 2 固定效应系数 γ 的边缘后验分布 $f(\gamma | Y)$ 通过积分求得:

$$f(\gamma | Y) = \iiint f(\beta, \gamma, \sigma_\varepsilon^2, T | Y) \partial \beta \partial \sigma_\varepsilon^2 \partial T$$

则可以根据前面的讨论进行假设检验。类似地,有层 1 随机效应系数 β 和层 1 残差方差 σ_ε^2 的边缘后验分布:

$$f(\beta | Y) = \iiint f(\beta, \gamma, \sigma_\varepsilon^2, T | Y) \partial \gamma \partial \sigma_\varepsilon^2 \partial T$$

$$f(\sigma_\varepsilon^2 | Y) = \iiint f(\beta, \gamma, \sigma_\varepsilon^2, T | Y) \partial \beta \partial \gamma \partial T$$

由此可以依据这些后验分布对 β 、 γ 、 σ_ε^2 等进行假设检验。

3 数据来源与基本分析

3.1 数据来源

从中国环境监测总站爬取到成渝地区双城经济圈 2015—12—2022—12 期间, $PM_{2.5}$ 、 NO_2 、 O_3 、 SO_2 、 CO 五项常规污染物的逐时监测数据,共计 50 万 1 千多条。这些污染物数据涉及成渝地区双城经济圈 163 个国控空气质量监测站点,其中四川省有 123 个,分布于成都市、巴中市等 21 个市(州);重庆市有 40 个,分布于 9 个中心城区及 12 个周边区县。

3.2 基本分析

为更好地评估成渝地区双城经济圈近年来大气污染治理成效,先将各监测站点的 $PM_{2.5}$ 逐时监测数据汇总平均计算出年平均浓度,可视化分析成渝两地的雾霾浓度变化轨迹;再根据 $PM_{2.5}$ 浓度大于 $150 \mu g/m^3$ 为重度及以上污染状态的空气质量等级划分标准,统计出 2015—2022 年成渝两地 163 个监测站点所属 42 个地区的空气质量处于重度及以上污染状态的时间比例,对比分析各地区空气质量变化差异(图 1)。

(1) 年度空气质量趋势分析。根据图 1(a),2014 年除 1414A 站点外,其他站点的 $PM_{2.5}$ 浓度均介于 $50 \sim 60 \mu g/m^3$;2021—2022 年所有站点浓度都在 $40 \mu g/m^3$ 以下;2016 年、2019 年和 2021 年均有小幅度上升;可见重庆市变化趋势整体呈下降态势,但有波动,空气质量不太稳定。从 1(b) 基线密集程度不难发现:2014 年大部分站点 $PM_{2.5}$ 浓度集中在 $45 \sim 70 \mu g/m^3$;而后阶段即 2022 年, $PM_{2.5}$ 浓度集中于 $20 \sim 50 \mu g/m^3$,呈现前分后聚整体下降趋势。由此猜想:四川省大气污染治理成效明显,但基线错综复杂,如站点 2561A、1887A、1899A 波动较大,站点 2561A 除 2018 年外, $PM_{2.5}$ 年均浓度值均在 $35 \mu g/m^3$ 标准下,但 2018 年超过了 $40 \mu g/m^3$;站点 1887A 整体呈下降趋势,但 2017 年有略微上升,2019 年上升幅度较大;站点 1899A 更是呈波浪式变化。

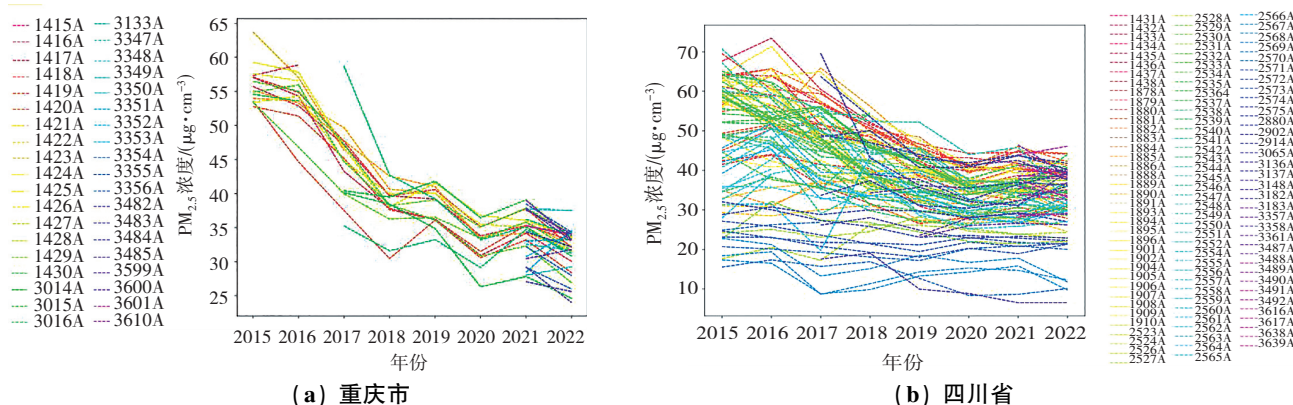


图 1 重庆市、四川省各监测站点 PM_{2.5} 年均浓度变化趋势图

Fig. 1 Annual average concentration of PM_{2.5} at monitoring stations in Chongqing City and Sichuan Province

(2) 监测站点所属地区污染状态时间占比分析。表 1 为成渝地区 2015—2022 年重度及以上污染状态地区占比。根据表 1, 2015—2022 年, 除四川省 3 州外, 成渝两地各地区年年都陷入“霾伏”, 总体上四川省大气污染比重庆市严重, 重污染区域主要为成都平原、川南、川东北地区, 其中成都、德阳、自贡、宜宾、达州等是污染最严重的城市。重度及以上污染时间比率从 2015 年比率最大值四川省自贡市的 9.5 下降到 2022 年比率最大值四川省泸州市的 0.8, 可见城市空气质量总体呈

转好趋势, 空气治理成效显著。但大气污染综合治理并不平衡、稳定、协同, 诸多地区空气质量反弹较大。可将各地区 PM_{2.5} 浓度变化状态划分为两种类型, 一是 2015—2022 年呈逐年下降状态且变化显著, 如渝中区; 二是虽整体下降, 但偶有反弹, 属这一变化类型的城市居多。如成都市由 2015 年的最高值降低到 2019 年的最低值, 又在 2019—2022 年间波动; 达州市的占比从 2015 年开始持续下降, 但排名不断波动; 泸州前几年都较好, 在 2022 年却成为空气质量垫底城市。

表 1 成渝地区 2015—2022 年重度及以上污染状态地区占比

Table 1 Percentages of areas with severe or above pollution from 2015 to 2022 in Chengdu-Chongqing area %

2015	2016	2017	2018	2019	2020	2021	2022
自贡 9.5	自贡 7.5	自贡 16.8	自贡 3.4	达州 2.0	宜宾 2.5	宜宾 1.4	泸州 0.8
成都 6.1	成都 4.5	成都 5.2	绵阳 2.2	宜宾 1.8	达州 0.9	成都 0.9	自贡 0.6
内江 5.6	达州 3.6	宜宾 4.6	达州 2.2	乐山 1.0	自贡 0.9	泸州 0.8	宜宾 0.6
达州 4.9	泸州 3.5	乐山 4.4	宜宾 2.0	自贡 0.8	凉山州 0.8	达州 0.7	南充 0.5
眉山 4.8	宜宾 3.2	德阳 3.9	德阳 1.9	德阳 0.6	德阳 0.6	自贡 0.7	永川 0.5
泸州 4.7	德阳 2.9	渝中 3.8	成都 1.9	南充 0.6	成都 0.6	德阳 0.6	合川 0.5
渝中 4.7	眉山 2.3	泸州 3.5	乐山 1.2	成都 0.5	乐山 0.6	铜梁 0.5	乐山 0.4
宜宾 4.5	绵阳 1.8	眉山 3.5	广安 0.9	渝中 0.5	广安 0.3	南充 0.5	成都 0.4
巴南 4.4	巴南 1.7	达州 3.2	渝中 0.8	北碚 0.4	南充 0.2	绵阳 0.3	广安 0.3
渝北 4.3	乐山 1.6	内江 3.0	泸州 0.8	雅安 0.4	九龙坡 0.2	乐山 0.3	巴南 0.3
九龙坡 4.1	渝中 1.6	雅安 2.6	雅安 0.8	九龙坡 0.3	绵阳 0.2	广安 0.3	眉山 0.3
南岸 4.0	九龙坡 1.6	绵阳 2.6	内江 0.6	内江 0.3	北碚 0.2	南川 0.3	凉山州 0.2
沙坪坝 3.7	沙坪坝 1.5	渝北 2.0	九龙坡 0.6	泸州 0.3	巴南 0.2	内江 0.3	潼南 0.2
广安 3.5	内江 1.5	九龙坡 2.0	巴南 0.6	广安 0.2	内江 0.2	涪陵 0.3	德阳 0.2
德阳 3.5	渝北 1.3	南岸 2.0	眉山 0.5	绵阳 0.2	江北 0.2	合川 0.3	九龙坡 0.2
北碚 3.4	巴中 1.2	巴南 1.9	南岸 0.5	资阳 0.2	渝中 0.2	荣昌 0.2	铜梁 0.2
南充 3.1	资阳 1.2	沙坪坝 1.4	南充 0.5	巴中 0.2	眉山 0.1	眉山 0.2	达州 0.2
乐山 3.0	广安 1.1	北碚 1.2	资阳 0.5	南岸 0.2	资阳 0.1	资阳 0.2	绵阳 0.2
江北 2.9	南岸 1.1	广安 0.9	遂宁 0.4	渝北 0.2	泸州 0.1	攀枝花 0.2	万州 0.2
遂宁 1.8	南充 1.1	江北 0.9	渝北 0.4	江北 0.2	南岸 0.1	万州 0.2	荣昌 0.2
绵阳 1.6	雅安 1.0	资阳 0.8	北碚 0.4	沙坪坝 0.2	渝北 0.1	九龙坡 0.2	渝中 0.1
巴中 1.2	北碚 0.9	遂宁 0.7	沙坪坝 0.4	遂宁 0.1	攀枝花 0.1	遂宁 0.2	雅安 0.1
资阳 1.1	遂宁 0.8	南充 0.7	巴中 0.3	眉山 0.1	雅安 0.1	永川 0.1	内江 0.1
雅安 0.4	江北 0.7	巴中 0.5	江北 0.2	广元 0.1	沙坪坝 0.1	潼南 0.1	丰都县 0.1
攀枝花 0.1	广元 0.2	广元 0.1	广元 0.1	巴南 0.1	巴中 0.0	长寿 0.1	沙坪坝 0.1
凉山州 0.1	凉山州 0.0	攀枝花 0.1	攀枝花 0.0	攀枝花 0.0	广元 0.0	巴中 0.1	长寿 0.1
广元 0.1	攀枝花 0.0	甘孜州 0.1	甘孜州 0.0	凉山州 0.0	甘孜州 0.0	北碚 0.1	南岸 0.1
甘孜州 0.1	阿坝州 0.0	阿坝州 0.0	凉山州 0.0	甘孜州 0.0	遂宁 0.0	渝中 0.1	南充 0.1
阿坝州 0.0	甘孜州 0.0	凉山州 0.0	阿坝州 0.0	阿坝州 0.0	阿坝州 0.0	凉山州 0.1	江北 0.1

注:2021 年新增 15 个站点(乐山市、巴中市、泸州市、德阳市、广元市各新增 1 个,成都市新增 6 个,绵阳市、自贡市各新增 2 个),故 2021—2022 年共有 42 个地区纳入比较, 29 名后的未列在表中, 它们的占比均不超过 0.1。

综上,成渝地区双城经济圈近年来空气质量改善明显,整体上 $PM_{2.5}$ 年均浓度呈逐年下降趋势,但监测站点的个体变化差异大,各地区污染水平和改善效果不一致,站点 $PM_{2.5}$ 浓度初始水平和下降轨迹也存在差异。

4 雾霾监测层次贝叶斯模型建构

4.1 数据结构及指标体系

利用重度及以上污染状态时间占比这一指标说明地区雾霾污染的变化差异还不够充分,需构建模型对其影响因素做进一步地探索。首先,以 $PM_{2.5}$ 浓度为因变量;其次,影响变量选取 4 个监测指标: NO_2 、 O_3 、 SO_2 、 CO ,以及从各地区统计年鉴获取的气象、人口、经济、产业结构、农业因素等指标。整理监测数据和统计年鉴年度指标数据,该数据具有显著的“年度时间-站点-地区”三层嵌套结构特征,即年度时间嵌套于站点,站点嵌套于地区。由此构建了年度时间-站点-地区三水平雾霾监测影响因素指标体系和发展模型。由于模型构建中剔除了不显著变量,故表 2 的指标体系仅展示各层级的显著变量。

表 2 年度时间-站点-地区三水平雾霾监测影响因素指标体系

Table 2 Index system of influencing factors of annual time-site-region three-level haze monitoring

层级	变量表示	代表指标
因变量	y	$PM_{2.5}(\mu g/m^3)$
水平 1	time	年度时间,0-6 分别表示 2015—2021 年
水平 2	A1	$CO(\mu g/m^3)$
	A2	$SO_2(\mu g/m^3)$
	A3	$O_3(\mu g/m^3)$
	A4	$NO_2(\mu g/m^3)$
水平 3	B1	平均气温($^{\circ}C$)
	B2	日照时数(h)
	B3	平均相对湿度(%)
	B5	人口密度(人/ km^2)
	B7	城镇化率(%)
	B8	全社会固定资产(万元)
	B10	地区生产总值(万元)
B15	第二产业产值占 GDP 比重(%)	
B17	农用化肥施用量(t)	

4.2 三水平发展模型建构

由基本分析看到:各监测站点 $PM_{2.5}$ 浓度虽整体呈下降趋势,但各站点间的变化差异大,波动速率不一,显然不适合利用传统线性回归模型。建立层次模型,从监测站点和地区不同层面找到解释变量反映这种差异更科学。

(1) 建立空模型。首先建立空模型,检验数据是否是分层嵌套数据。根据表 3,三水平方差均显著,来自水平 1 时间的方差为 77.295,方差占比 31.29%;来

自水平 2 站点的方差为 15.971,方差占比为 6.46%;来自水平 3 地区的方差为 153.790,方差占比为 62.25%。水平 3 和水平 2 的方差是组间方差,组内相关系数为 $R_{ICC} = 68.71\%$ 。可见所有站点 $PM_{2.5}$ 初始水平显著不同,有 68.71% 的变异是由地区和站点差异引起的,故应考虑构建三水平层次模型进行分析。

表 3 $PM_{2.5}$ 浓度影响因素空模型估计结果

Table 3 Null model estimation results of $PM_{2.5}$ concentration influencing factors

	参数	估计值	标准误	P 值
固定效应	β_0	44.909	2.596	<0.001
	$\sigma_{\mu_0}^2$ (水平 3)	153.790	46.671	0.001
随机效应	$\sigma_{\mu_0}^2$ (水平 2)	15.971	5.164	0.002
	σ_{ϵ}^2 (水平 1)	77.295	5.424	<0.001

(2) 建立三水平无条件发展模型。在考虑各层次存在差异的基础上,继续建立具有随机系数的无条件模型,估计结果见表 4。表 4 中最后三列为没有层级结构的 OLS 估计,用于对比 OLS 估计和 IGLS 估计效果。从参数估计显著性角度来看:固定效应参数显著,说明站点 $PM_{2.5}$ 平均浓度初始值(2015 年)为 $50.441 \mu g/m^3$,2015—2021 年显著降低,且每年平均降低 $2.792 \mu g/m^3$;随机效应参数中,三水平上斜率参数和截距参数的协方差并不显著,其余方差参数都显著,说明发展模型的截距和斜率在站点和地区之间存在明显差异,即 $PM_{2.5}$ 平均浓度随时间的变化因站点和地区不同而有差异。模型整体拟合效果方面,层次模型的 IGLS 估计与线性回归模型的 OLS 估计结果非常接近,但层次模型的 $-2\ln(\text{likelihood})$ 和 AIC 明显小于一般线性回归模型,说明层次发展模型的拟合效果更好。

(3) 建立多变量三水平发展模型。三水平无条件发展模型表明 $PM_{2.5}$ 平均浓度初始水平不仅因监测站点而异,且随时间的变化率也不同,即发展模型的截距和时间斜率在监测站点之间有明显差异。为继续探索引起 β_{0ij} 和 β_{1ij} 变化的因素,构建多变量三水平发展模型,估计结果见表 5。首先,从固定效应参数显著性角度看,变量 ZB7、ZB10 不论是 IGLS 估计还是 OLS 估计均不显著;变量 ZB1、ZB5、ZB8、ZB15 的多层次模型 IGLS 估计不显著,而一般线性回归模型 OLS 估计显著;时间斜率 β_1 和 ZA1、ZA2、ZA4 以及 ZA1 与时间的交互作用($ZA1 * t$)系数的层次模型 IGLS 估计显著,而线性回归模型 OLS 估计不显著。其次,从随机效应参数的显著性角度看,除 $\sigma_{\mu_0}^2$ 外,多水平发展模型的方差参数估计均显著。这应值得注意,在空模型中此系数式显著,在多变量条件模型中反而不显著,说明加入影响

β_{0ij} 和 β_{1ij} 的解释变量十分有价值。最后,从模型整体拟合效果来看,三水平发展模型的 $-2\ln(\text{likelihood})$ 和 AIC 均明显小于一般线性回归模型的 OLS 估计。由此可见,层次模型的拟合效果明显优于一般线性模型。

表 4 无条件三水平发展模型估计结果

Table 4 Estimation results of unconditional three-level development model

参 数	IGLS 估计			OLS 估计			
	估计值	标准误	P 值	估计值	标准误	P 值	
固定效应	β_0	50.441	2.824	<0.001	51.100	1.149	<0.001
	β_1 (水平 1)	-2.792	0.505	<0.001	-2.840	0.471	<0.001
	$\sigma_{\mu 0}^2$ (水平 3)	178.966	55.192	0.001			
随机效应	$\sigma_{\nu 1}^2$ (水平 3)	4.940	1.756	0.005			
	$\sigma_{\nu 01}$ (水平 3)	-11.355	7.496	0.130			
	$\sigma_{\mu 0}^2$ (水平 2)	21.581	4.992	<0.001			
	σ_e^2 (水平 1)	46.232	3.346	<0.001	224.331	14.076	<0.001
	$-2\ln(\text{likelihood})$		3 625.374			4 191.508	
AIC		3 639.374			4 197.508		

表 5 多变量三水平发展模型参数估计结果

Table 5 Parameter estimation results of multivariate three-level development model

参 数	IGLS 估计			OLS 估计			
	估计值	标准误	P 值	估计值	标准误	P 值	
β_0	48.791	1.978	<0.001	47.885	0.919	<0.001	
β_1 (水平 1)	-1.098	0.567	0.053	-0.597	0.404	0.139	
ZA1(水平 2)	1.422	0.730	0.051	1.297	0.902	0.150	
ZA2(水平 2)	2.309	0.743	0.006	0.830	0.911	0.362	
ZA3(水平 2)	-1.633	0.442	<0.001	-2.660	0.521	<0.001	
ZA4(水平 2)	0.600	0.321	0.062	0.741	0.495	0.134	
ZA1 * t(水平 2)	0.801	0.359	0.026	0.531	0.442	0.230	
ZA2 * t(水平 2)	1.384	0.369	<0.001	1.817	0.432	<0.001	
ZB1(水平 3)	-0.143	0.670	0.831	1.026	0.605	0.090	
ZB2(水平 3)	-2.615	1.211	0.031	-6.640	0.982	<0.001	
ZB3(水平 3)	3.374	1.327	0.011	3.736	1.041	<0.001	
ZB5(水平 3)	0.390	2.329	0.867	2.324	1.157	0.045	
ZB7(水平 3)	3.008	2.921	0.303	-0.812	1.310	0.536	
ZB8(水平 3)	-0.044	0.565	0.938	-1.170	0.635	0.066	
ZB10(水平 3)	0.243	7.176	0.973	3.200	2.317	0.167	
ZB15(水平 3)	-0.702	0.734	0.339	1.934	0.613	0.002	
ZB17(水平 3)	4.438	1.737	0.011	2.978	0.648	<0.001	
ZB7 * t(水平 3)	-0.759	0.432	0.079	-0.703	0.341	0.039	
随机效应	$\sigma_{\nu 0}^2$ (水平 3)	69.696	22.489	0.398			
	$\sigma_{\nu 1}^2$ (水平 3)	3.769	1.340	0.002			
	$\sigma_{\nu 01}$ (水平 3)	-3.393	4.018	0.005			
	$\sigma_{\mu 0}^2$ (水平 2)	12.102	3.107	<0.001			
	σ_e^2 (水平 1)	34.354	2.512	<0.001	99.057	6.265	<0.001
$-2\ln(\text{likelihood})$		3 395.809			3 716.784		
AIC		3 342.809			3 756.784		

注:指标都经标准化处理,标准化后均添加了 Z,如 ZA1 表示 CO($\mu\text{g}/\text{m}^3$)的标准化。

4.3 三水平发展模型贝叶斯分析

由于多变量三水平发展模型的 IGLS 估计方法对变量 $ZB1$ 、 $ZB5$ 、 $ZB7$ 、 $ZB8$ 、 $ZB10$ 、 $ZB15$ 的估计不显著,但这些变量却与 $PM_{2.5}$ 存在相关性,故对多变量三水平发展模型尝试使用贝叶斯推断估计,对比分析贝叶斯估计方法是否使模型估计效果提高;同时将经典 IGLS 估计、经验贝叶斯和完全贝叶斯方法对比,论证完全贝叶斯方法的优势,估计结果如表 6 所示。在完全贝叶斯方法中,考虑到先验信息获取的难易程度,将固定效应参数和随机效应参数的先验分布都设为正态先验这一常用基本分布。在经验贝叶斯估计方法中,将水平 3

中的重要变量 $ZB7$ 、 $ZB8$ 、 $ZB10$ 的先验分布设置为正态先验,其他参数采用 IGLS 估计。

根据表 6,首先从参数估计的显著性来看,变量 $ZB1$ 、 $ZB5$ 、 $ZB15$ 和水平 3 斜率和截距的协方差 $\sigma_{\mu 01}$ 在经验贝叶斯方法的估计下不显著,而在完全贝叶斯方法的估计下显著。其次从参数估计的精度即标准误差角度来看,完全贝叶斯方法的标准误差无论是固定效应参数还是随机效应参数都比经验贝叶斯方法的标准误差小。最后从模型整体拟合效果来看,完全贝叶斯方法的 DIC 为 3 191.731,小于经验贝叶斯的 3 200.338。上述均说明完全贝叶斯方法优于经验贝叶斯方法。

表 6 多变量三水平发展模型的贝叶斯估计结果
Table 6 Bayesian estimation results of multivariate three-level development model

参 数	IGLS 估计		经验贝叶斯估计		完全贝叶斯估计	
	估计值(标准误)	P 值	估计值(标准误)	P 值	估计值(标准误)	P 值
β_0	48.791(1.978)	<0.001	49.298(2.084)	<0.001	48.955(0.471)	<0.001
β_1 (水平 1)	-1.098(0.567)	0.053	-1.085(0.541)	0.045	-1.075(0.328)	0.001
ZA1(水平 2)	1.422(0.730)	0.051	1.439(0.751)	0.055	1.419(0.387)	<0.001
ZA2(水平 2)	2.309(0.743)	0.006	2.044(0.753)	0.007	2.035(0.402)	<0.001
ZA3(水平 2)	-1.633(0.442)	<0.001	-1.685(0.453)	<0.001	-1.645(0.323)	<0.001
ZA4(水平 2)	0.600(0.321)	0.062	0.601(0.320)	0.060	0.610(0.260)	0.019
ZA1 * t(水平 2)	0.801(0.359)	0.026	0.830(0.361)	0.022	0.820(0.258)	0.001
ZA2 * t(水平 2)	1.384(0.369)	<0.001	1.350(0.372)	<0.001	1.364(0.247)	<0.001
ZB1(水平 3)	-0.143(0.670)	0.831	-0.161(0.674)	0.811	-0.148(0.050)	0.003
ZB2(水平 3)	-2.615(1.211)	0.031	-2.494(1.272)	0.050	-2.535(0.452)	<0.001
ZB3(水平 3)	3.374(1.327)	0.011	3.159(1.374)	0.022	3.070(0.457)	<0.001
ZB5(水平 3)	0.390(2.329)	0.867	0.108(1.602)	0.946	0.200(0.050)	<0.001
ZB7(水平 3)	3.008(2.921)	0.303	3.000(0.050)	<0.001	3.000(0.050)	<0.001
ZB8(水平 3)	-0.044(0.565)	0.938	-0.044(0.005)	<0.001	-0.044(0.005)	<0.001
ZB10(水平 3)	0.243(7.176)	0.973	0.298(0.049)	<0.001	0.298(0.049)	<0.001
ZB15(水平 3)	0.702(0.734)	0.339	0.779(0.706)	0.270	0.703(0.193)	<0.001
ZB17(水平 3)	4.438(1.737)	0.011	4.119(1.643)	0.012	4.024(0.481)	<0.001
ZB7 * t(水平 3)	-0.759(0.432)	0.079	-0.821(0.469)	0.080	-0.773(0.308)	0.012
$\sigma_{\mu 0}^2$ (水平 3)	69.696(22.489)	0.398	101.970(40.170)	0.011	70.059(3.051)	<0.001
$\sigma_{\mu 1}^2$ (水平 3)	3.769(1.340)	0.002	5.081(1.945)	0.009	3.998(0.178)	<0.001
$\sigma_{\mu 01}$ (水平 3)	-3.393(4.018)	0.005	-5.037(6.151)	0.413	-3.010(0.519)	<0.001
$\sigma_{\mu 0}^2$ (水平 2)	12.102(3.107)	<0.001	12.352(3.425)	<0.001	12.012(0.529)	<0.001
σ_e^2 (水平 1)	34.354(2.512)	<0.001	35.302(2.651)	<0.001	34.915(1.309)	<0.001
-2ln(likelihood)	3 395.809		Deviance(MCMC)	3 200.338	3 191.731	
AIC	3 342.809		DIC	3 302.600	3 286.94	

4.4 模型解读及主要结论

成渝地区双城经济圈雾霾监测影响因素三水平贝叶斯发展模型分析结果表明:雾霾污染水平逐年下降,且地区间下降速率不一。其中,雾霾污染水平由大气

污染物、气象、人口、经济、产业结构、农业生产综合因素导致,雾霾污染水平下降速率受到 CO 、 SO_2 浓度以及城镇化率的影响。

具体来看,在监测站点层面上, $PM_{2.5}$ 浓度受 SO_2

浓度的影响最大,其次是CO,最后是NO₂;O₃浓度增加会抑制雾霾污染,CO、SO₂、NO₂浓度增加会引起PM_{2.5}浓度增加;此外,ZA1、ZA2与时间的交互效应显著,说明CO和SO₂浓度越高,PM_{2.5}浓度增长越快。在地区层面上,平均相对湿度、人口密度、城镇化率、地区生产总值、第二产业产值占GDP比重、农用化肥施用量对雾霾污染有正相关关系;平均气温、日照时数、全社会固定资产系数对雾霾污染有负相关关系。其中,农用化肥施用量对PM_{2.5}浓度的影响程度最大;城镇化率与PM_{2.5}浓度增长速度之间有负相关关系;ZB7与时间的交互效应显著,说明城镇化率与PM_{2.5}浓度增长速度之间有负相关关系。

值得注意的是, $\sigma_{i,01}$ 系数为负,说明地区这一层面上,PM_{2.5}初始浓度与变化速度总体呈负相关关系,这反映了PM_{2.5}浓度越高,雾霾治理越困难。

5 存在问题及对策建议

5.1 存在问题

大气污染联防联控有待加强。川渝两地空气质量总体为“气象影响型”,但两地的气象、生态环境部门尚未联合下发关于共同提升大气污染防治的政策文件,整体区域的大气污染协同治理政策也较少。

产业转型升级,工业互联网融合不够。川南四市(自贡、泸州、内江、宜宾)大气质量垫底成渝地区双城经济圈,这四市恰是四川的“东北地区”。虽《成渝经济区南部城市群发展规划(2014—2020年)》明确对推动老工业基地调整升级做出规划,但目前川南经济区大数据智能化牵引产业转型升级不够,“互联网+工业”仍未得到充分利用。

尚未形成有效的农业源大气污染防治机制。农用化肥中包含了形成大气主要污染物,尤其是形成臭氧污染因子的三大NH。但现行涉及农业面源污染的政策法规,特别是生态环境部编制的《成渝地区双城经济圈生态环境保护规划》,主要集中在水生态环境治理,对农业源大气污染防治不多,对化肥施用的监测及管理更少。

5.2 对策建议

深化多部门协同联动和跨区域联防联控机制。首先,深入贯彻落实碳达峰、碳中和战略部署和渝府办发〔2022〕22号文件精神,尽快构建新一代大气污染防治科学体系;其次,生态环境部门应当同气象局、农业农村部等多部门协同开展基于气温、相对湿度、日照时数等气象监测和农业生产中化肥、农药使用量等数据的大气污染变化趋势分析;利用监测、模拟以及卫星、航测、遥感等新技术分析成渝两地重点区域内大气污染来源及其变化趋势,落实成渝地区空气污染常态化精

准监测;最后,注重各城市产业结构等特征条件,实施差异化治理、精细化治理和动态化治理,针对各城市制定差异化碳减排和污染物减排目标,以及防止污染状况起伏波动,采取动态化治理方式,实现双城经济圈持续绿色低碳发展。

强化技术支撑推动区域产业结构绿色转型升级。第一,推动传统行业绿色化改造。成都、重庆要发挥“双核引领”作用,充分借助物联网、大数据、人工智能等新技术,全面推进大渡口、永川、自贡等老工业基地钢铁、建材、天然气、化工等传统领域企业实施全要素、全流程清洁化、循环化、低碳化改造。第二,培育绿色新兴产业集群。推进传统产业集群和开发区整合,围绕新一代新能源、新材料、新能源汽车、集成电路、智能网联汽车等新兴产业,并着力为产业厚植“智能因子”,打造能源资源消耗低,污染物排放量小,附加值高,对川渝协同发展有支撑力的绿色低碳优势产业集群。第三,重点培育氢能产业生态圈。围绕“成渝氢走廊”政策机遇,内江市、资阳市等周边城市鼎力培育氢能产业生态圈,各城市联动合作,发展制氢、储氢、加氢、用氢等全产业链技术,发展清洁能源科技,推动氢能应用,推动成渝地区实现氢能物流、氢能交通、氢能工业、氢能建筑等。

重点加强工业源SO₂和农业源氨污染防治。在全面执行控制挥发性有机物(VOCs)排放限值的同时,SO₂排放源和农用化肥污染源也应引起高度重视。要强化重点行业污染治理,加快火电、钢铁、水泥、焦化及燃煤工业锅炉超低排放改造和使用高污染燃料的燃料类工业炉窑清洁能源替代。将工业污染源的全面达标排放计划持续推进,在线实时监测大气污染物,建立实时流数据库,对未达标的企业依法进行停产限产整治。引导和监督农民开展种养结合,实现畜禽粪肥还田利用,减少化肥施用;优化施肥模式,普及化肥减量增效技术,集成推广化肥施用绿色防控模式。

参考文献(References):

- [1] 童英华,冯忠岭,张占莹.基于AHP的雾霾影响因素评价分析[J].西南师范大学学报(自然科学版),2020,45(3):87—94.
TONG Ying-hua, FENG Zhong-ling, ZHANG Zhan-ying. On evaluation for influential factors of fog-haze based on AHP[J]. Journal of Southwest China Normal University (Natural Science Edition), 2020, 45(3): 87—94.
- [2] 谭秉霖,欧阳强.湖南省城市雾霾污染特征及影响因素分析[J].衡阳师范学院学报,2020,41(6):108—115.
TAN Bing-lin, OUYANG Qiang. An analysis of characteristics of urban haze pollution and its impact factors in Hunan province[J]. Journal of Hengyang Normal University, 2020,

- 41(6): 108—115.
- [3] 许昌日, 朱法华, 刘丹丹, 等. 基于污染物和气象要素的北京市雾霾影响因素分析[J]. 电力科技与环保, 2021, 37(1): 1—8.
XU Chang-ri, ZHU Fa-hua, LIU Dan-dan, et al. Analysis of the influencing factors of haze in Beijing based on pollutants and meteorological elements [J]. Electric Power Technology and Environmental Protection, 2021, 37(1): 1—8.
- [4] 陈弄祺, 许瀛. 北京雾霾污染影响因素实证分析[J]. 中国人口·资源与环境, 2016, 26(S2): 73—76.
CHEN Nong-qi, XU Ying. Empirical analysis on the influencing factors of haze pollution in Beijing [J]. China Population, Resources and Environment, 2016, 26 (S2): 73—76.
- [5] 王艳, 段学军. 氨污染: 被忽视的雾霾元凶[J]. 生态经济, 2017, 33(6): 6—9.
WANG Yan, DUAN Xue-jun. Ammonia pollution: The neglected culprit of smog[J]. Ecological Economy, 2017, 33(6): 6—9.
- [6] 高明, 张晨璐. 三大城市群雾霾污染影响因素实证对比分析[J]. 太原理工大学学报(社会科学版), 2022, 40(1): 77—84.
GAO Ming, ZHANG Chen-lu. Empirical comparative analysis of influencing factors of haze pollution in three city clusters [J]. Journal of Taiyuan University of Technology (Social Science Edition), 2022, 40(1): 77—84.
- [7] 任毅, 郭丰. 长江经济带雾霾污染空间效应与影响因素分析[J]. 经济与管理, 2019, 33(1): 43—50.
REN Yi, GUO Feng. Analysis of the spatial effects and influencing factors of haze pollution on the Yangtze River economic belt [J]. Economy and Management, 2019, 33(1): 43—50.
- [8] 肖翠翠, 周景博, 郑晨昱. 雾霾的区域性特征及空间关联关系研究: 以辽中南城市群为例[J]. 东北大学学报(社会科学版), 2021, 23(3): 22—29.
XIAO Cui-cui, ZHOU Jing-bo, ZHENG Chen-yu. Study on the regional characteristics and spatial relationship of haze: the case of central and southern Liaoning urban agglomeration [J]. Journal of Northeastern University (Social Science Edition), 2021, 23(3): 22—29.
- [9] 盛小星, 叶春明. 基于集对分析法的长三角雾霾风险评估[J]. 资源开发与市场, 2017, 33(3): 334—337, 359.
SHENG Xiao-xing, YE Chun-ming. Haze risk assessment based on set pair analysis in Changjiang River delta [J]. Resource Development & Market, 2017, 33(3): 334—337, 359.
- [10] 洪宇. 基于结构方程模型的我国区域雾霾水平评价[J]. 统计与决策, 2019, 35(2): 62—65.
HONG Yu. Evaluation of regional smog level in China based on structural equation model [J]. Statistics & Decision, 2019, 35(2): 62—65.
- [11] 王霞. 基于改进灰色关联度的郑州市雾霾影响因素分析[J]. 数学的实践与认识, 2020, 50(1): 143—153.
WANG Xia. Smog factor analysis of Zhengzhou based on the improved grey correlation analysis [J]. Mathematics in Practice and Theory, 2020, 50(1): 143—153.
- [12] 孙亚男, 费锦华. 基于机器学习的雾霾污染精准治理[J]. 资源科学, 2021, 43(5): 872—885.
SUN Ya-nan, FEI Jin-hua. Precise governance of haze pollution based on machine learning [J]. Resources Science, 2021, 43(5): 872—885.
- [13] 沈晓晶, 姜祝辉, 陈建, 等. 黄渤海晴空像元下雾霾识别的机器学习算法[J]. 遥测遥控, 2021, 42(3): 74—78.
SHEN Xiao-jing, JIANG Zhu-hui, CHEN Jian, et al. Machine learning algorithm of haze detection over the Yellow and Bohai Sea [J]. Journal of Telemetry, Tracking and Command, 2021, 42(3): 74—78.
- [14] 张敏, 姜锐, 李勇, 等. 基于高层次结构数据的多水平模型贝叶斯推断及应用[J]. 数理统计与管理, 2019, 38(2): 357—366.
ZHANG Min, JIANG Rui, LI Yong, et al. Bayesian inference and its applications for multilevel model with high-level structural data [J]. Journal of Applied Statistics and Management, 2019, 38(2): 357—366.
- [15] 张敏, 李勇. 基于层次模型的全国主要城市雾霾监测评估及影响因素研究[J]. 数理统计与管理, 2022, 41(4): 587—598.
ZHANG Min, LI Yong. Research on smog monitoring and evaluation and influencing factors in major cities of China based on hierarchical model [J]. Journal of Applied Statistics and Management, 2022, 41(4): 587—598.
- [16] (美) CLARK J S. 面向生态学数据的贝叶斯统计: 层次模型、算法和 R 编程 [M]. 沈泽昊, 等, 译. 北京: 科学出版社, 2013.
CLARK J S. Bayesian Statistics for Ecological data models for ecological data [M]. Shen Ze-hao, et al. Trans. Beijing: Science Press, 2013.
- [17] (美) 斯蒂芬 W 劳登布什, (美) 安东尼·S. 布里克. 分层线性模型: 应用与数据分析方法 [M]. 郭志刚, 等, 译. 2 版. 北京: 社会科学文献出版社, 2016.
LOUDENBUSH S W, ANTHONY S. Bryk Hierarchical linear models [M]. GUO Zhi-gang, et al. Trans. 2nd ed. Beijing: Social Sciences Literature Publishing House, 2016.

责任编辑: 李翠薇